Choosing a neighbourhood in Toronto, Canada

EJ Romero

2021May30

Background

- The city of Toronto has 140 neighbourhoods that is a viable location for a certain stakeholder/s,
- A single family wants to potentially start a Thai restaurant business at and live nearby.
- Tasked to evaluate neighbourhoods and provide assessment based on health and safety factors for them.

Audience

- The target audience (or stakeholders) are mainly for individuals (in this case single-type families
- 1-4 members who want to start a food restaurant and move in the Toronto area.
- Use-case and dataset is also viable for individuals who want to start an ethnic-based food restaurant
- Focus on advising stakeholder/s wanting to open Asian/Pacific Islander cuisine.

Data

- Provided by the City of Toronto's Open Data Platform
    - Demographics
    - Economics
    - Environment
    - Health
    - Housing
    - Crime Rates
    - Safety
    - Total Population

Data Resources

- City of Toronto's Open Data Platform: https://open.toronto.ca/
- Toronto Police: https://data.torontopolice.on.ca/
- Neighbourhood Crime Rates 2011: https://open.toronto.ca/dataset/neighbourhood-crime-rates/
- Neighbourhood Crime Rates 2020: https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-2020-1
- Toronto Environment: https://open.toronto.ca/dataset/wellbeing-toronto-environment/
- Toronto Health: https://open.toronto.ca/dataset/wellbeing-toronto-health/
- Toronto Safety: https://open.toronto.ca/dataset/wellbeing-toronto-safety/
- Toronto Housing: https://open.toronto.ca/dataset/wellbeing-toronto-housing/
- Toronto Economics: https://open.toronto.ca/dataset/wellbeing-toronto-economics/
- Toronto Demographics: https://open.toronto.ca/dataset/wellbeing-toronto-demographics/
- Canada Inflation Rate: https://tradingeconomics.com/canada/inflation-cpi

Data Notice

- Majority of Open Data is from 2011
- 2016 and 2020 data available on some datasets
  - Will be included as updated data
- Features (columns) per dataset may contain zero values
- Zero values will remain in assessment
  - Will be removed before choosing the best neighbourhood

Data Pre-processing

- Data processing evaluated using Jupyter Notebook and Microsoft Excel.
- 140 neighbourhoods (observations) per dataset and different features.
- Assessment based on only 40 neighbourhoods

```
New Toronto dataframe shape: (40, 6)
<class 'pandas.core.frame.DataFrame'>
```

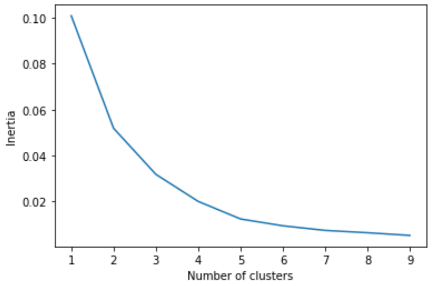
	Cluster Labels	Neighbourhood Id	Neighbourhood	Borough	Postal Code	Latitude	Longitude
62	1	63	The Beaches	East Toronto	M4E	43.676357	-79.293031
63	1	64	Woodbine Corridor	East Toronto	M4E	43.676357	-79.293031
64	1	65	Greenwood-Coxwell	East Toronto	M4L	43.668999	-79.315572
65	1	66	Danforth	East Toronto	M4C	43.695344	-79.318389
66	4	67	Playter Estates-Danforth	East Toronto	M4K	43.679557	-79.352188
67	4	68	North Riverdale	East Toronto	M4K	43.679557	-79.352188
68	4	69	Blake-Jones	East Toronto	M4J	43.685347	-79.338106
69	4	70	South Riverdale	East Toronto	M4K	43.679557	-79.352188

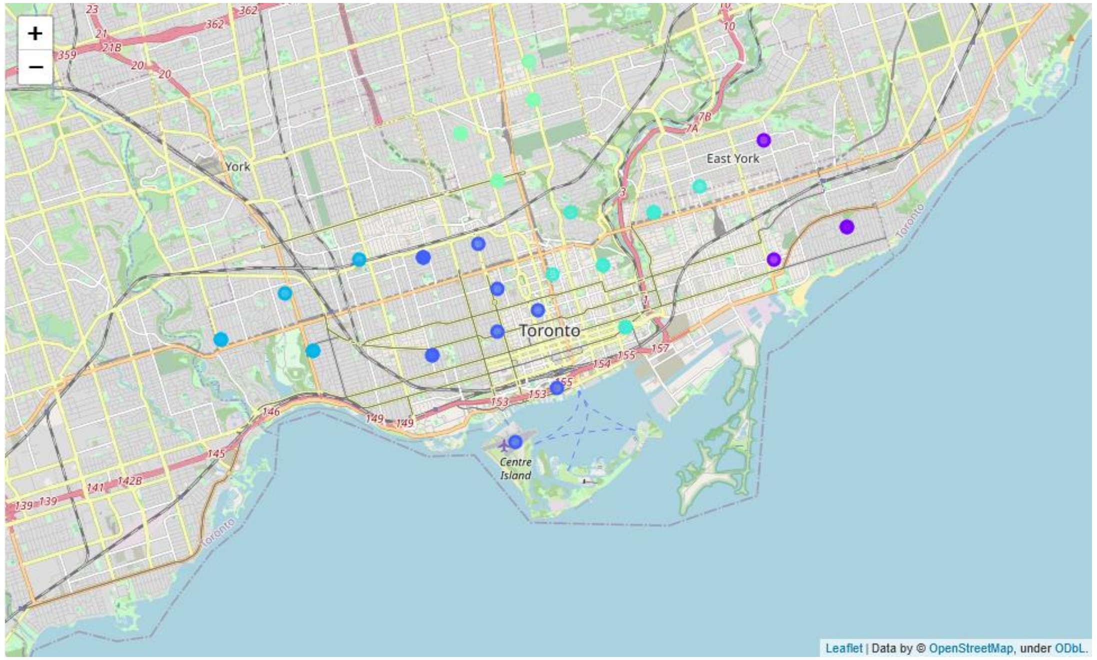
KMeans Clustering

```python
# Select appropriate cluster number
choose_k = toronto_area.drop(['Neighbourhood Id', 'Neighbourhood', 'Postal Code', 'Borough'], 1)

inertias = []

for i in range(1, 10):
    kmeans = KMeans(n_clusters=i, random_state=0)
    kmeans.fit(choose_k)
    inertias.append(kmeans.inertia_)

plt.plot(range(1, 10), inertias)
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
```

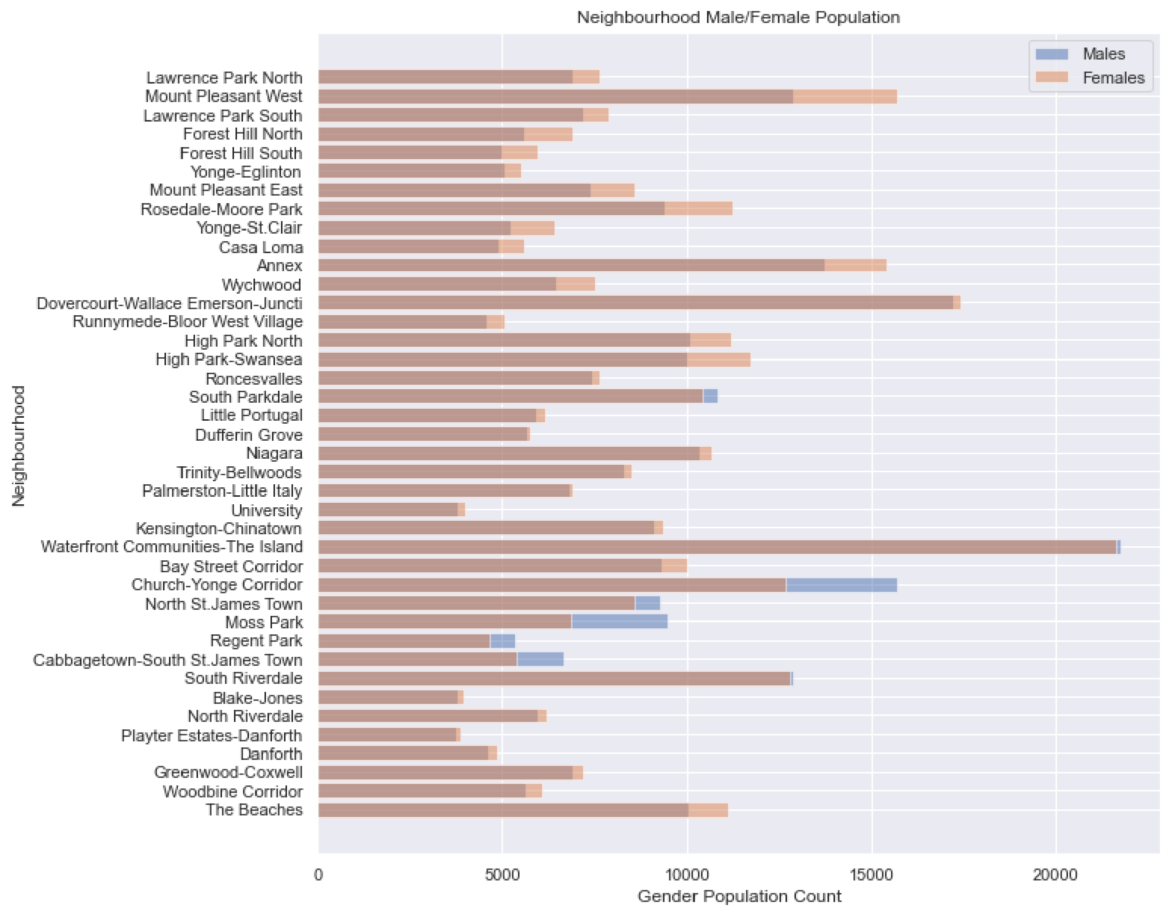
```
C:\anaconda\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:881: UserWarning: KMeans is known to have a memory lea
k on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment var
iable OMP_NUM_THREADS=1.
  warnings.warn(
```

```
Text(0, 0.5, 'Inertia')
```





Data Exploration 1.0

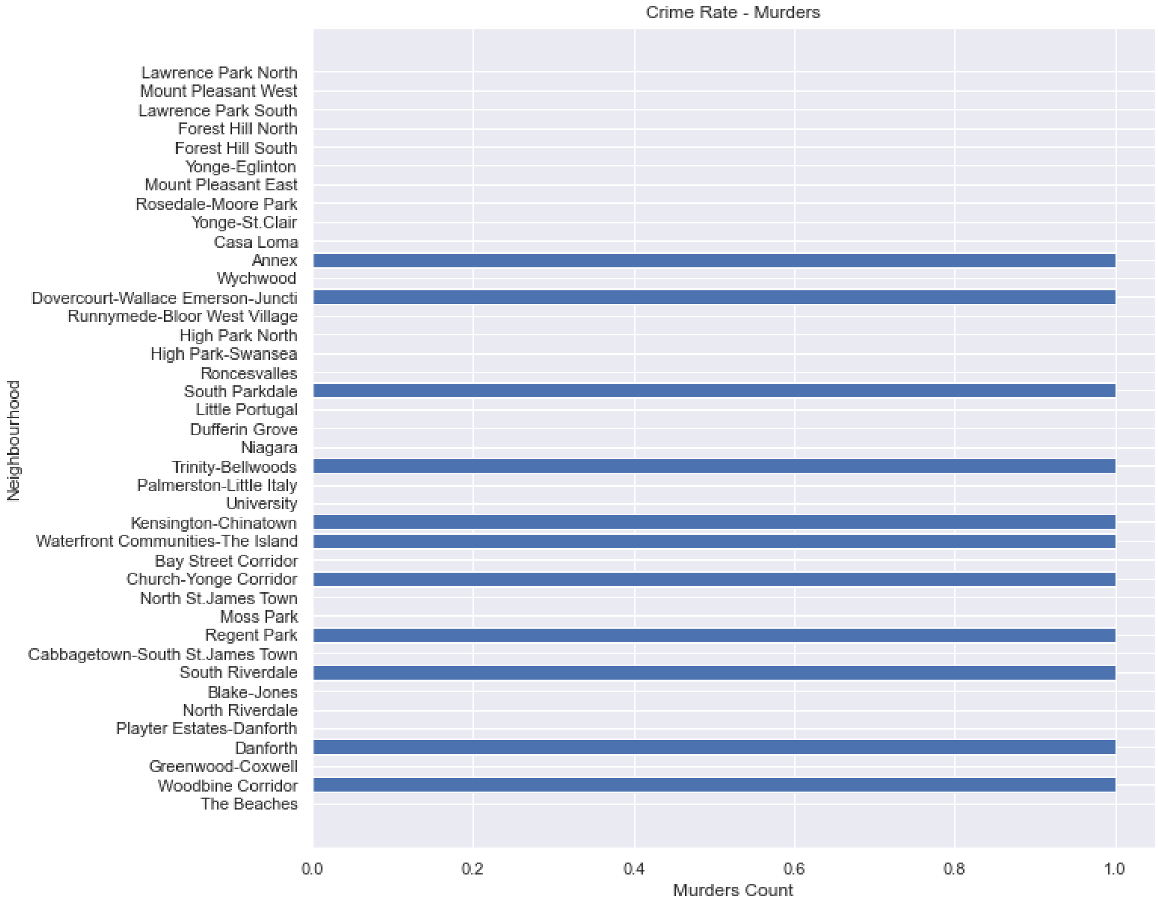


Data Exploration 1.0 (Continued)

```python
# Check where maximum and minimum provided Asian languages are located
lang_chi_max = sub_demo_df_3[sub_demo_df_3['   Language - Chinese'] == sub_demo_df_3['   Language - Chinese'].max()]
print(lang_chi_max[['Neighbourhood', '   Language - Chinese']])
lang_chi_min = sub_demo_df_3[sub_demo_df_3['   Language - Chinese'] == sub_demo_df_3['   Language - Chinese'].min()]
print(lang_chi_min[['Neighbourhood', '   Language - Chinese']])
print()

lang_kor_max = sub_demo_df_3[sub_demo_df_3['   Language - Korean'] == sub_demo_df_3['   Language - Korean'].max()]
print(lang_kor_max[['Neighbourhood', '   Language - Korean']])
lang_kor_min = sub_demo_df_3[sub_demo_df_3['   Language - Korean'] == sub_demo_df_3['   Language - Korean'].min()]
print(lang_kor_min[['Neighbourhood', '   Language - Korean']])
print()

lang_tag_max = sub_demo_df_3[sub_demo_df_3['   Language - Tagalog'] == sub_demo_df_3['   Language - Tagalog'].max()]
print(lang_tag_max[['Neighbourhood', '   Language - Tagalog']])
lang_tag_min = sub_demo_df_3[sub_demo_df_3['   Language - Tagalog'] == sub_demo_df_3['   Language - Tagalog'].min()]
print(lang_tag_min[['Neighbourhood', '   Language - Tagalog']])
print()

lang_tam_max = sub_demo_df_3[sub_demo_df_3['   Language - Tamil'] == sub_demo_df_3['   Language - Tamil'].max()]
print(lang_tam_max[['Neighbourhood', '   Language - Tamil']])
lang_tam_min = sub_demo_df_3[sub_demo_df_3['   Language - Tamil'] == sub_demo_df_3['   Language - Tamil'].min()]
print(lang_tam_min[['Neighbourhood', '   Language - Tamil']])
```

```
          Neighbourhood     Language - Chinese
77  Kensington-Chinatown                  6070
        Neighbourhood     Language - Chinese
100  Forest Hill South                   180

         Neighbourhood     Language - Korean
75  Bay Street Corridor                815
   Neighbourhood     Language - Korean
68   Blake-Jones                  20

         Neighbourhood     Language - Tagalog
73  North St.James Town                 1560
   Neighbourhood     Language - Tagalog
78    University                   40

         Neighbourhood     Language - Tamil
73  North St.James Town               915
              Neighbourhood     Language - Tamil
79   Palmerston-Little Italy                 0
100        Forest Hill South                 0
```

Data Exploration 1.0 (Continued)



Data Exploration 2.0

- Safety Factors
- Health Factors
- Keep zero values are kept
  - Casa Loma
- Maximum, Mean, Minimum techniques

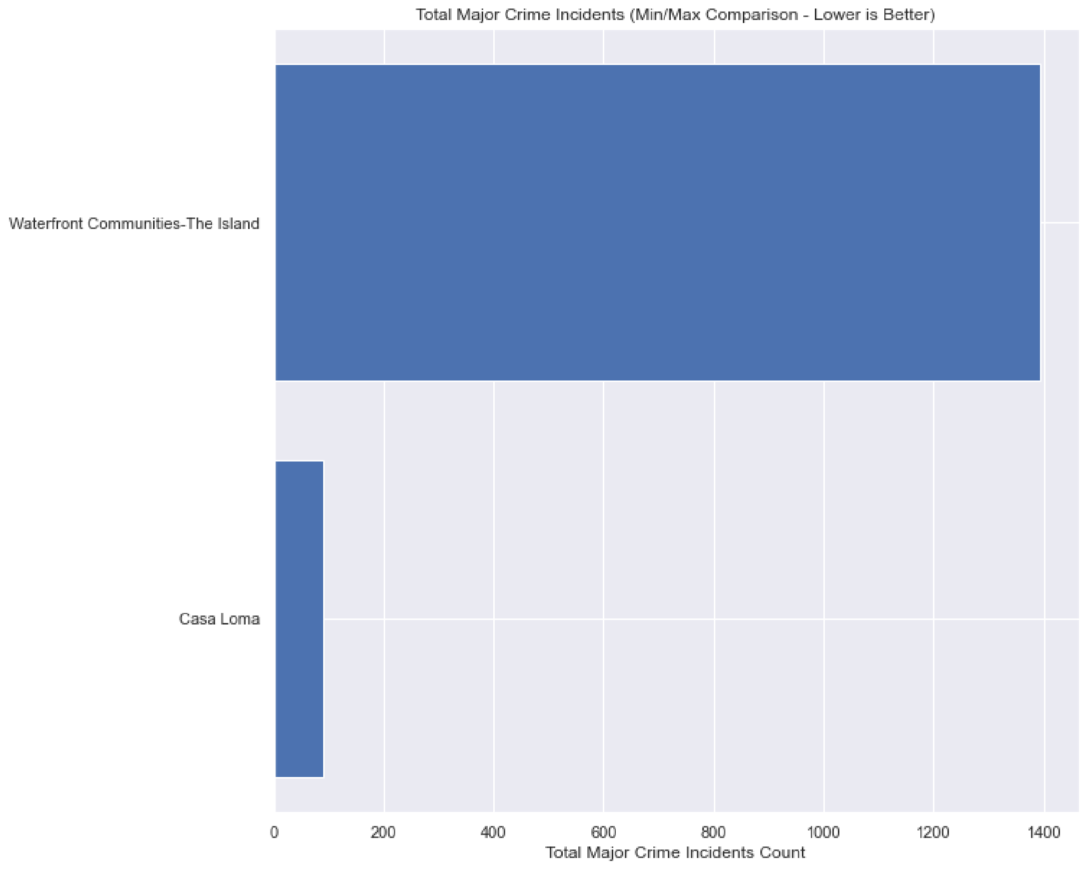
```
# Neighbourhood with the least amount of combined incidents
safety_combo = tor_sub_merge_2011_test[['Neighbourhood', 'Total Major Crime Incidents',
                                        'Fires & Fire Alarms', 'Hazardous Incidents']]
safety_combo_2 = safety_combo[(safety_combo['Total Major Crime Incidents']<=350)
             & (safety_combo['Fires & Fire Alarms']<=250)
             & (safety_combo['Hazardous Incidents']<=150)]
```

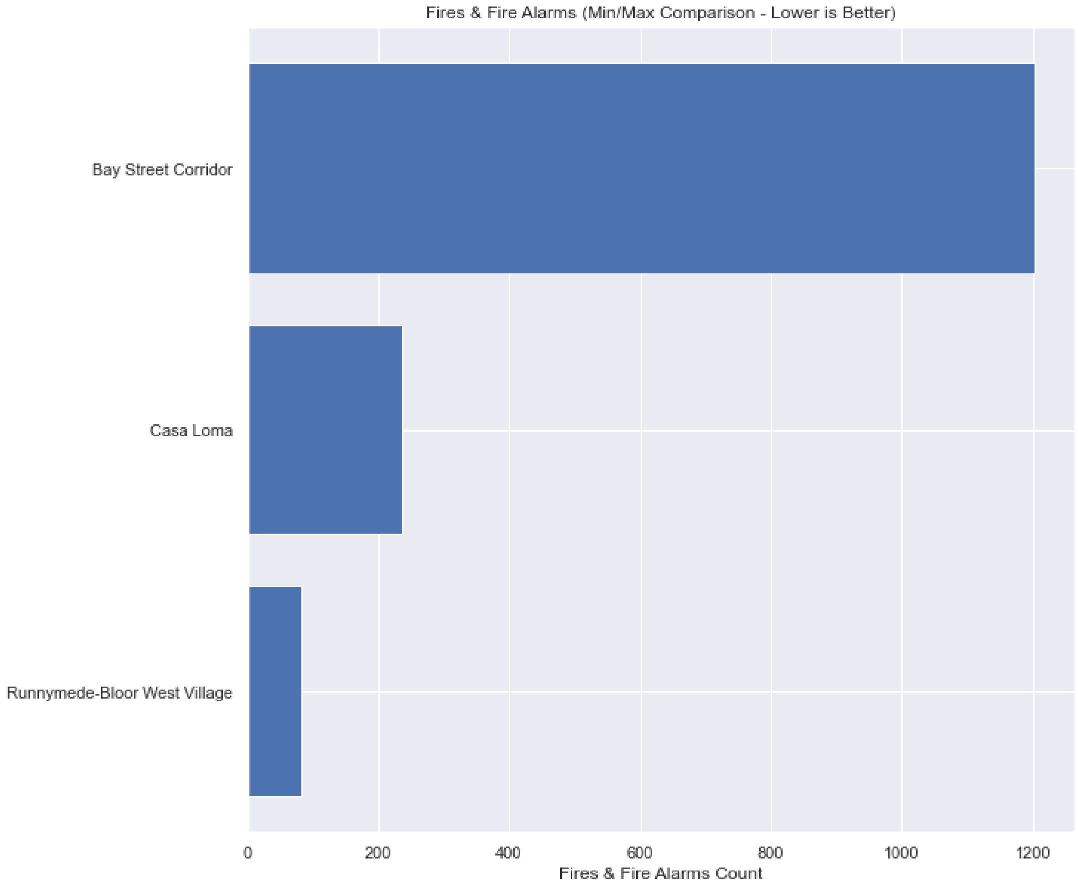
```
# Neighbourhood with of combined health perks
health_combo = tor_sub_merge_2011_test[['Neighbourhood', 'Pollutant Carcinogenic TEP Score',
                                        'Pollutants Released to Air', 'Health Providers']]
health_combo_2 = health_combo[(health_combo['Pollutant Carcinogenic TEP Score']<=100)
             & (health_combo['Pollutants Released to Air']<=1500)
             & (health_combo['Health Providers']>=50)]
```

```
# Merge (inner join) the two combo subset dataframes
toronto_cluster_filtered = toronto_cluster.merge(safety_health_combo, how='inner',
                                                 left_on='Neighbourhood', right_on='Neighbourhood')
toronto_cluster_filtered
```

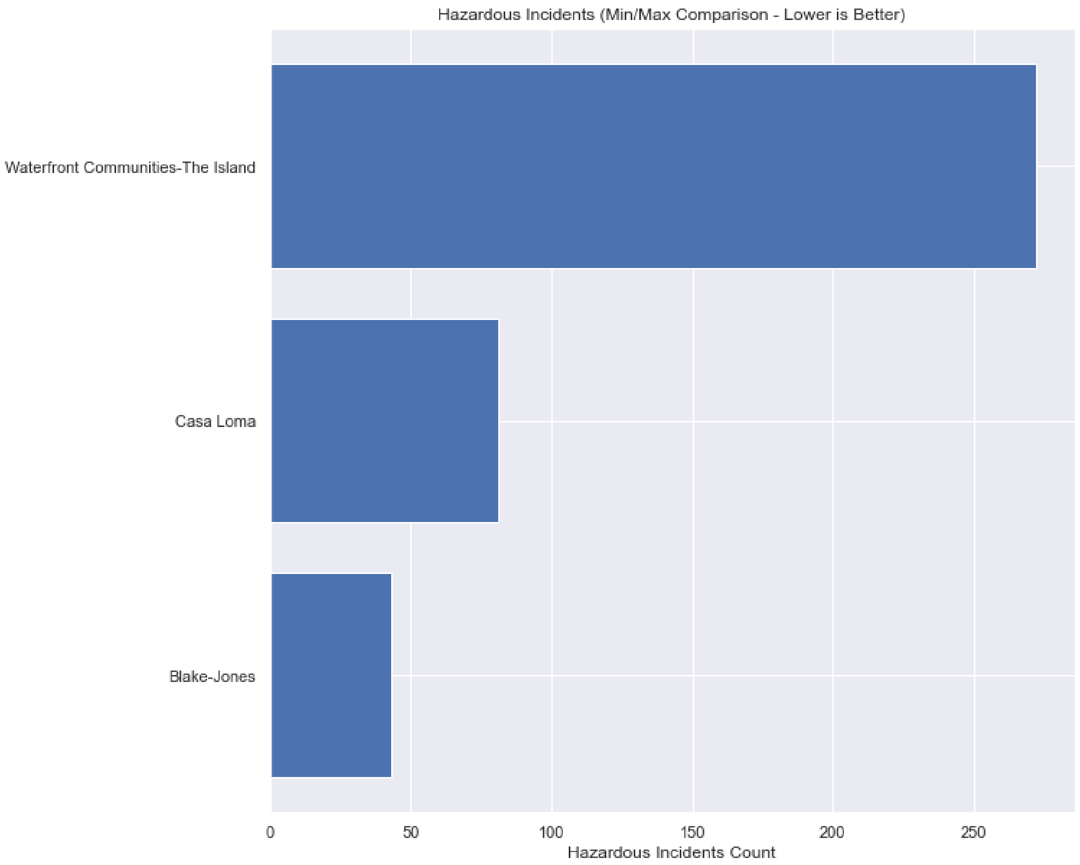
	Cluster Labels	Neighbourhood Id	Neighbourhood	Borough	Postal Code	Latitude	Longitude	Total Major Crime Incidents	Fires & Fire Alarms	Hazardous Incidents	Pollutant Carcinogenic TEP Score	Pollutants Released to Air	Health Providers
0	1	66	Danforth	East Toronto	M4C	43.695344	-79.318389	262	119	72	0.00	0	53
1	2	80	Palmerston-Little Italy	Downtown Toronto	M6G	43.669542	-79.422564	261	149	99	0.00	0	57
2	0	96	Casa Loma	Central Toronto	M4V	43.686412	-79.400049	91	236	81	75.84	575	77
3	0	97	Yonge-St.Clair	Central Toronto	M4V	43.686412	-79.400049	111	175	67	0.00	0	56
4	0	100	Yonge-Eglinton	Central Toronto	M4P	43.712751	-79.390197	229	147	115	0.00	0	62

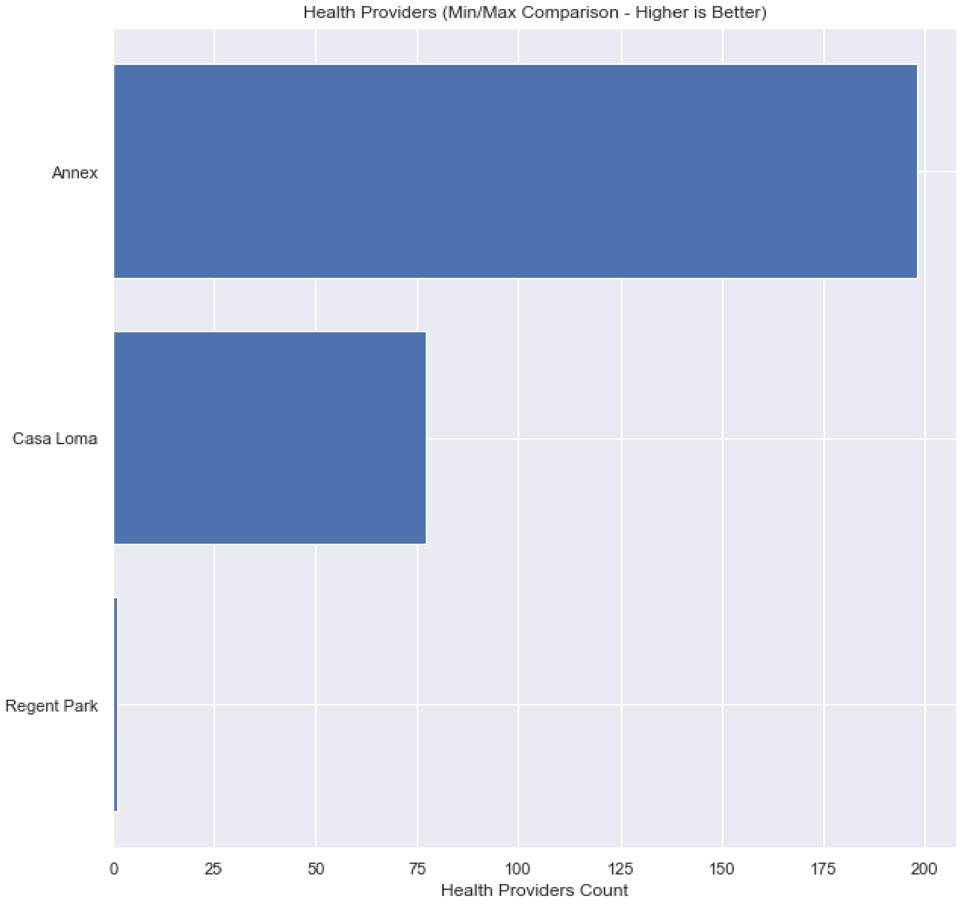
Data Exploration 2.0 (Continued)



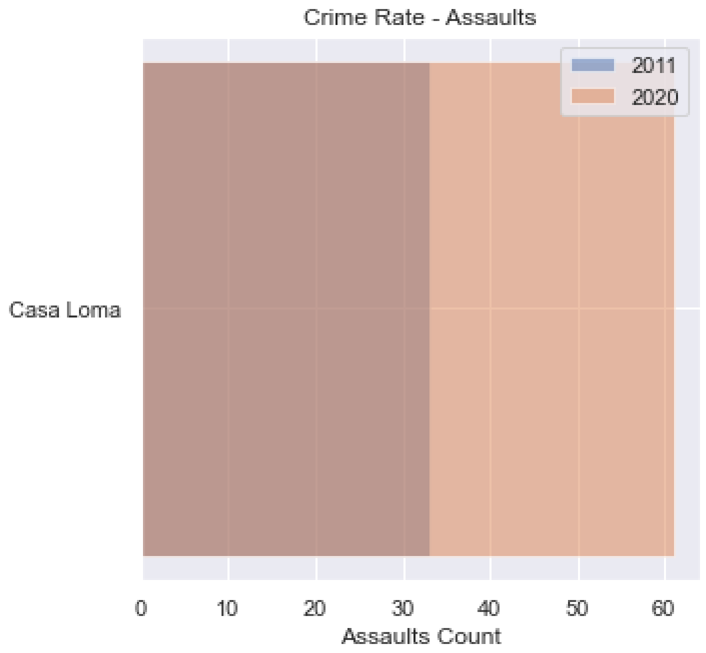


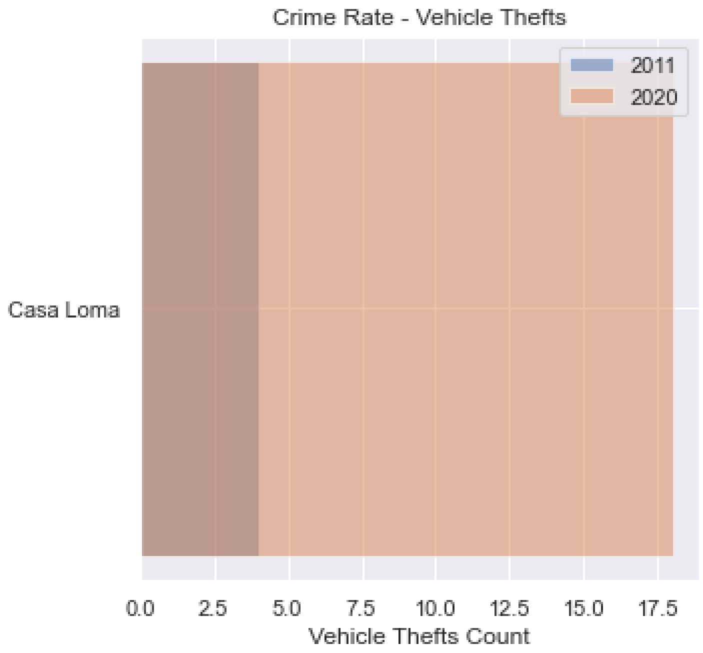
Data Exploration 2.0 (Continued)



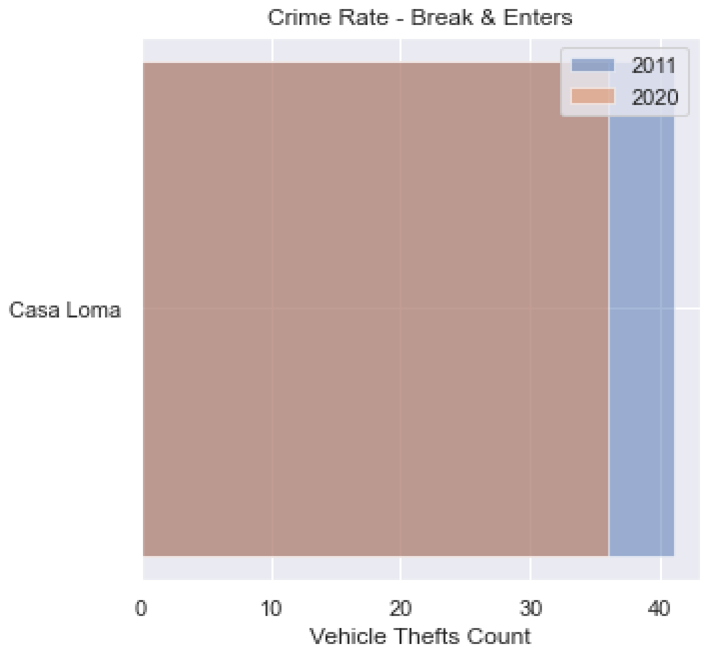


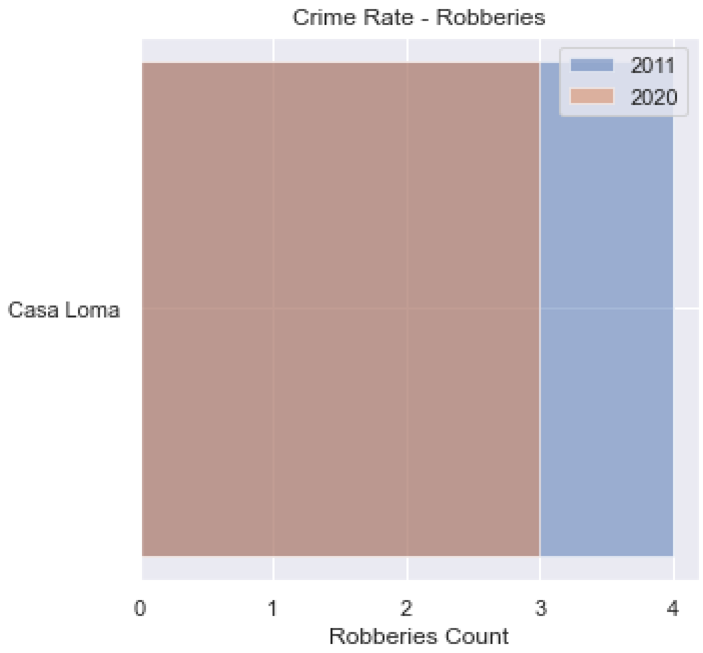
Data Exploration 2.0 (Continued)



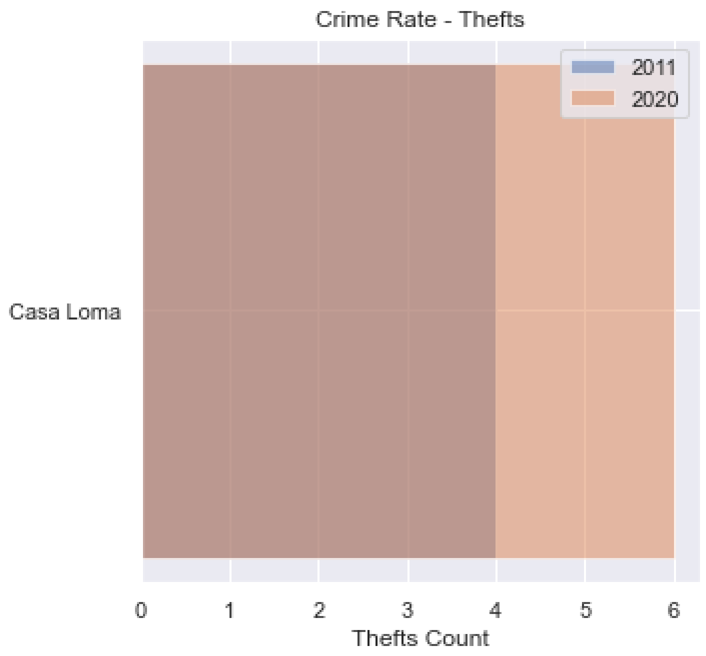


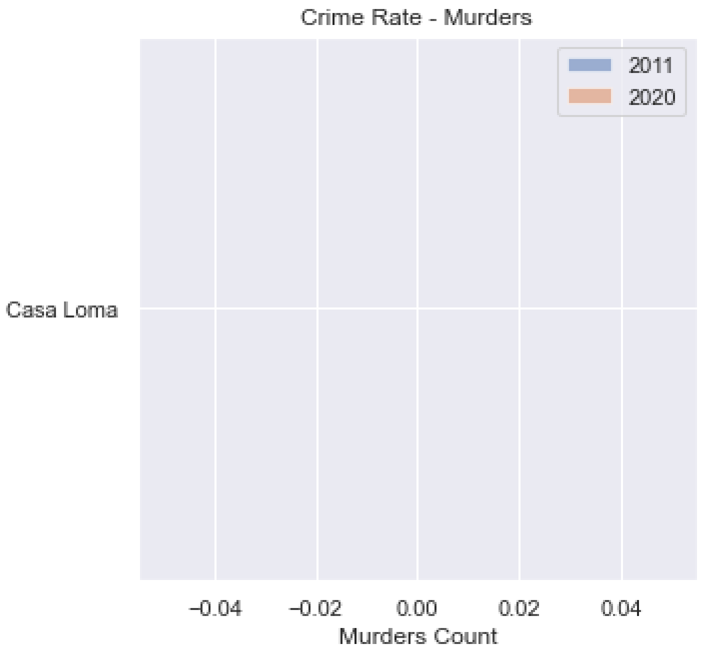
Data Exploration 2.0 (Continued)





Data Exploration 2.0 (Continued)

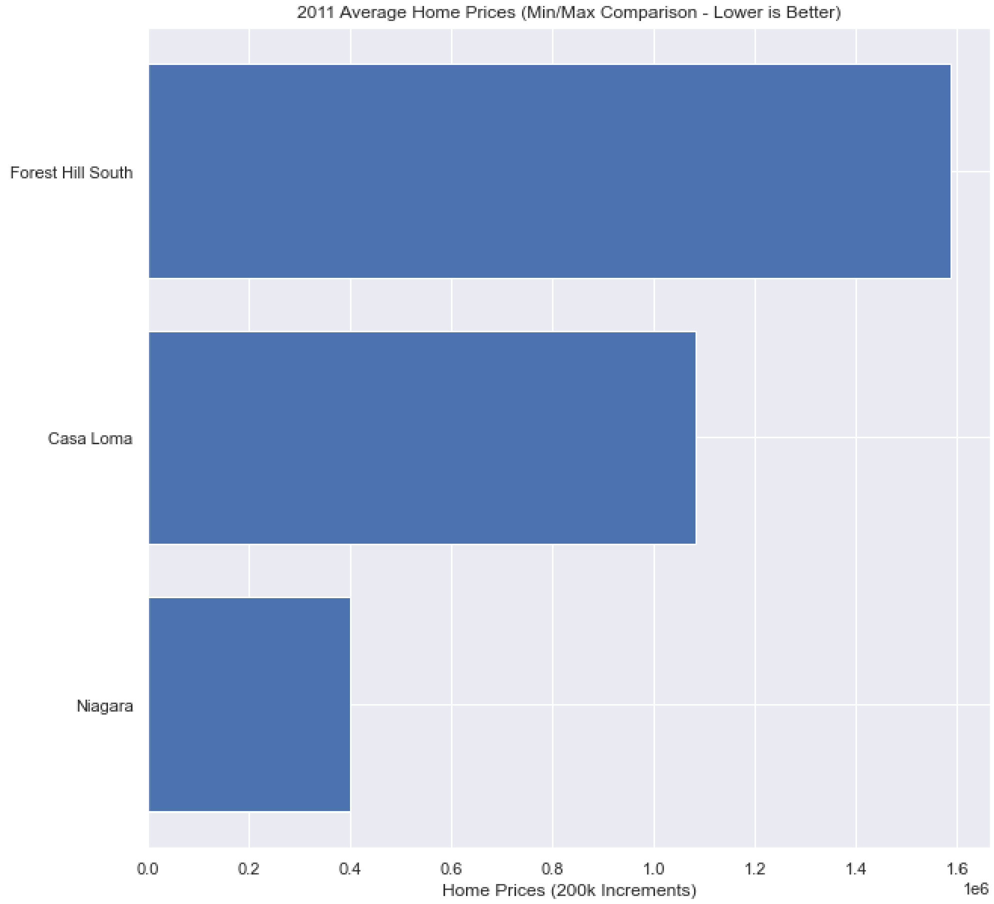




Data Exploration 2.0 (Continued)

```
# Find Min/Mean/Max home prices in 2011
home_price_max = tor_sub_merge_2011_test[tor_sub_merge_2011_test['Home Prices'] ==
                                        tor_sub_merge_2011_test['Home Prices'].max()]
home_price_mean = tor_sub_merge_2011_test['Home Prices'].mean()
home_price_min = tor_sub_merge_2011_test[tor_sub_merge_2011_test['Home Prices'] ==
                                        tor_sub_merge_2011_test['Home Prices'].min()]

print(home_price_max[['Neighbourhood', 'Home Prices']])
print()
print('2011 Average Home Prices: {}'.format(home_price_mean))
print()
print(home_price_min[['Neighbourhood', 'Home Prices']])
```

```
        Neighbourhood  Home Prices
35  Forest Hill South      1585984

2011 Average Home Prices: 702095.1

   Neighbourhood  Home Prices
19       Niagara       398281
```



Foursquare Data Assessment

```
toronto_neighborhood.iloc[2]
```

```
Cluster Labels                                    0
Neighbourhood Id                                 96
Neighbourhood                             Casa Loma
Borough                             Central Toronto
Postal Code                                     M4V
Latitude                                  43.686412
Longitude                                -79.400049
Total Major Crime Incidents                      91
Fires & Fire Alarms                             236
Hazardous Incidents                              81
Pollutant Carcinogenic TEP Score              75.84
Pollutants Released to Air                      575
Health Providers                                 77
Name: 2, dtype: object
```

```python
# Specify coordinates
neighborhood_latitude = toronto_neighborhood.loc[2, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = toronto_neighborhood.loc[2, 'Longitude'] # neighborhood longitude value

neighborhood_name = toronto_neighborhood.loc[2, 'Neighbourhood'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_latitude,
                                                               neighborhood_longitude))
```

```
Latitude and longitude values of Casa Loma are 43.6864123, -79.4000493.
```

```python
search_query = 'Thai'
radius = 100
print(search_query + ' .... OK!')
```

```
Thai .... OK!
```

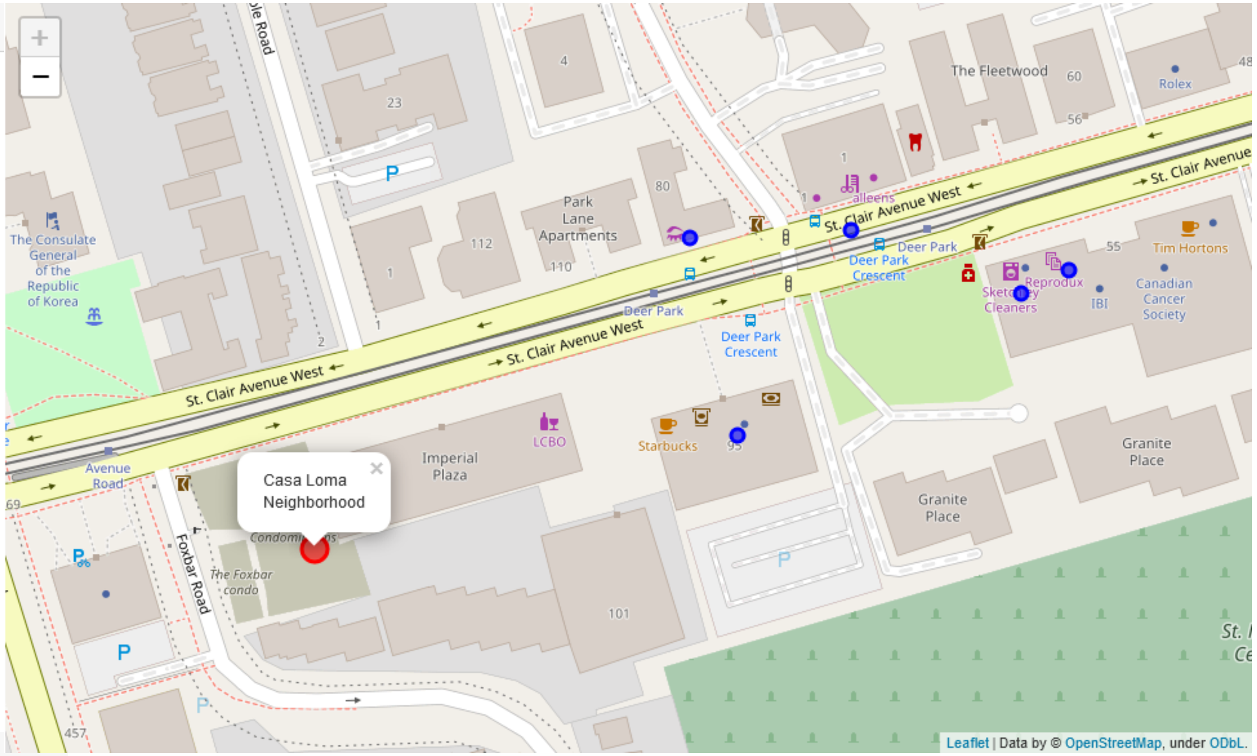
```python
# Query Foursquare
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '60b31bd1f3fd206e117773d8'},
 'response': {'venues': [{'id': '5a67afb973fe2528841f60f3',
    'name': 'The Market By Longo's',
    'location': {'address': '111 St Clair Ave W',
     'lat': 43.686711,
     'lng': -79.399536,
     'labeledLatLngs': [{'label': 'display',
       'lat': 43.686711,
       'lng': -79.399536}],
     'distance': 53,
     'postalCode': 'M4V 1N5',
     'cc': 'CA',
     'city': 'Toronto',
     'state': 'ON',
     'country': 'Canada',
     'formattedAddress': ['111 St Clair Ave W',
      'Toronto ON M4V 1N5',
      'Canada']},
    'categories': [{'id': '52f2ab2ebcbc57f1066b8b46',
```

Foursquare Data Assessment (Continued)

```
# Check values in the "categories" column
dataframe_filtered['categories'].value_counts()
```

```
Office                                     5
Building                                   3
Light Rail Station                         3
Residential Building (Apartment / Condo)   3
Embassy / Consulate                        2
Government Building                        2
Pharmacy                                   2
Cemetery                                   1
Doctor's Office                            1
Park                                       1
Fabric Shop                                1
Coffee Shop                                1
Dog Run                                    1
Advertising Agency                         1
Dentist's Office                           1
Café                                       1
Afghan Restaurant                          1
Spiritual Center                           1
Spa                                        1
Elementary School                          1
Salon / Barbershop                         1
Liquor Store                               1
Bank                                       1
Assisted Living                            1
Diner                                      1
Other Great Outdoors                       1
Athletics & Sports                         1
Auditorium                                 1
Italian Restaurant                         1
College Rec Center                         1
Roof Deck                                  1
Supermarket                                1
Name: categories, dtype: int64
```




```
# Example - restaurant = dataframe_filtered[dataframe_filtered['categories'] == 'Restaurant']
restaurant_1 = dataframe_filtered[dataframe_filtered['categories'] == 'Coffee Shop']
restaurant_2 = dataframe_filtered[dataframe_filtered['categories'] == 'Café']
restaurant_3 = dataframe_filtered[dataframe_filtered['categories'] == 'Afghan Restaurant']
restaurant_4 = dataframe_filtered[dataframe_filtered['categories'] == 'Diner']
restaurant_5 = dataframe_filtered[dataframe_filtered['categories'] == 'Italian Restaurant']
```

Conclusion

- Identify potential neigbourhoods to start a Thai restaurant business
  - Health and Safety Factors
  - Home prices
- Update datasets when available
  - Data comparison
  - Check for data correlation